EMPLOYING VAGUE PRIOR INFORMATION IN THE CONSTRUCTION

OF CONFIDENCE SETS

George Casella and Jiunn Tzon Hwang
Cornell University

BU-778-M Revised February 1986

ABSTRACT

In the problem of estimating the mean, θ, of a multivariate normal distribution, an experimenter will often be able to give some vague prior specifications about θ. This information is used to construct confidence sets centered at improved estimators of θ. These sets are shown to have uniformly (in θ) higher coverage probability than the usual confidence set (a sphere centered at the observations), with no increase in volume. Further, through the use of a modified empirical Bayes argument, a variable radius confidence set is constructed which provides a uniform reduction of volume. Strong numerical evidence is presented which shows that the empirical Bayes set also dominates the usual confidence set in coverage probability. All these improved sets provide substantial gains if the prior information is correct. Also considered are extensions to the unknown variance case, and a discussion of applications to the one-way analysis of variance. In particular, a procedure is presented which uniformly improves upon Scheffé's method of estimation of contrasts.

Keywords: MULTIVARIATE NORMAL MEAN; STEIN ESTIMATION; EMPIRICAL BAYES; ANALYSIS OF VARIANCE

Technical Report BU-778-M in the Biometrics Unit

1. INTRODUCTION

In the past twenty years, much progress has been made in the problem of improving on the usual point estimator of a multivariate normal mean. However, only recently has there been progress on the problem of improving upon the usual confidence set. Let $X \sim N(\theta,\sigma^2 I)$ where, for now, σ^2 is assumed to be known. The usual confidence set is a p-dimensional sphere centered at X with radius $c\sigma$, i.e.,

$$C^0_{X,\sigma} = \{\theta: |\theta - X| \le c\sigma\} .$$

The constant c is chosen to satisfy $P(\chi^2_p \le c^2) = 1-\alpha$, which implies that $P(\theta \epsilon C^0_{X,\sigma}) = 1-\alpha$ and $C^0_{X,\sigma}$ has coverage probability $1-\alpha$.

The usual confidence set $C^0_{X,\sigma}$ is minimax in the sense that among all the confidence sets with coverage probability at least $1-\alpha$, $C^0_{X,\sigma}$ minimizes the maximum volume. Despite this optimality property (and many others), it has been shown independently by Brown (1966) and Joshi (1967) that $C^0_{X,\sigma}$ can be improved provided $p \ge 3$. They showed that if $p \ge 3$, there exists another confidence set C^{BJ} dominating $C^0_{X,\sigma}$ in the sense that

$$\text{(a)} \quad P_\theta(\theta \epsilon C^{BJ}) \ge P_\theta(\theta \epsilon C^0_{X,\sigma}) ,$$

$$\text{(b)} \quad \text{volume } C^{BJ} \le \text{volume } C^0_{X,\sigma} ,$$

with strict inequality holding in either (a) or (b) for a set of positive Lebesgue measure of θ or X, respectively.

More recently, Faith (1976) derived an alternative confidence set by considering a version of a Bayes credible set. Berger (1980) also developed alternative confidence sets. Starting with a prior that gives admissible minimax point estimators, Berger constructed confidence ellipsoids centered at the posterior mean and oriented by the posterior covariance matrix. Berger and Faith both presented convincing analytical and numerical evidence that their confidence sets dominate $C^0_{X,\sigma}$.

Hwang and Casella (1982), and Hwang and Casella (1984) consider simpler confidence sets; spheres recentered at the positive part James-Stein estimator. These confidence sets are shown to dominate $C^0_{X,\sigma}$. In particular, Hwang and Casella (1982) give the first analytical proof that their confidence sets dominate $C^0_{X,\sigma}$. Later, Hwang and Casella (1984) provide another, simpler proof which strengthens these domination results. These stronger results form a base on which the results of this paper are built.

Even though $C^0_{X,\sigma}$ can be improved uniformly, it is impossible to significantly improve on $C^0_{X,\sigma}$ everywhere. (This is due to the minimaxity of $C^0_{X,\sigma}$.) So far, all the improved confidence sets proposed yield significant improvement over $C^0_{X,\sigma}$ (either by increasing coverage probability or decreasing volume) only when θ and X are near a fixed point. Naturally, an experimenter would choose the fixed point to be the most likely value of θ (or the prior guess of θ) so that there is a good chance of realizing a substantial gain by using the improved confidence set.

In some situations, however, there may only be vague prior information concerning the most likely value of θ. In particular, it may be thought that θ lies in a linear subspace of the parameter space, perhaps described by the equation $H\theta = 0$, where H is a known matrix. (In the point estimation problem, Bock (1982a, 1982b) has many interesting results concerning these and other forms of vague prior information.)

One useful type of prior information, particularly in the analysis of variance, is the specification that the θ_i's are equal to a common, unknown value. (This is the null hypothesis in the one-way analysis of variance.) We interpret this type of vague prior information as stating that the θ_i are close to each other, but it is not clear what is the common likely value. In such situations, all of the above confidence sets may improve upon $C^0_{X,\sigma}$ only slightly, and one might as well use $C^0_{X,\sigma}$.

In this paper we consider confidence sets that are recentered at estimators of the form

$$\delta^A(X) = AX + \{1-[a\sigma^2/|(I-A)X|^2]\}^+[(I-A)X] \ , \tag{1.1}$$

estimators which shrink toward a linear subspace. We pay particular attention to the matrix $A = (1/p)\underset{\sim}{1}\underset{\sim}{1}'$, where $\underset{\sim}{1}$ is a $p\times 1$ vector of ones. The resulting estimator is

$$\delta^L(X) = \bar{x}\underset{\sim}{1} + [1-(a\sigma^2/|X-\bar{x}\underset{\sim}{1}|^2)]^+(X-\bar{x}\underset{\sim}{1}) \ , \tag{1.2}$$

where $\bar{x} = (1/p)\Sigma^p_{i=1}X_i$, which is the positive part version of the estimator first derived by Lindley (1962). This estimator shrinks toward the estimate of the common mean, $\bar{x}$, and it is well known that, as a point estimator, $\delta^L(X)$ dominates X (under sum of squared errors loss), provided $0 < a \le 2(p-3)$. It is also known that $\delta^L(X)$ yields significant improvement over X as long as $\Sigma(\theta_i - \bar{\theta})^2/\sigma^2$ is small, where $\bar{\theta} = (1/p)\Sigma\theta_i$. Therefore, $\delta^L(X)$ is a particularly pertinent point estimator when it is thought that the θ_i's are close to each other. The same is true of the confidence sets based on δ^L, as will be shown in this paper.

In Section 2, we prove that the confidence set recentered at δ^L (with radius $c\sigma$) dominates $C^0_{X,\sigma}$, for a certain range. Generalized confidence sets centered at an improved estimator shrinking toward an arbitrary linear subspace are also constructed. Applications to the one-way analysis of variance model, and other models, are discussed.

Section 3 develops (using an empirical Bayes approach) confidence sets centered at δ^L with radius uniformly smaller than the usual confidence set. Numerical evidence shows that the coverage probability of these sets is at least $1-\alpha$. When σ^2 is unknown, but an independent estimate, s^2, of σ^2 is available, we modify these sets by replacing σ^2 by its estimate. Numerical evidence also confirms the superiority of this adaptive empirical Bayes confidence set over the usual confidence set (based on X and s^2). Section 4 discusses applications of these results to the multiple comparisons problem. In particular, it is shown that Scheffé's procedure can be improved uniformly (in the sense that, for the same confidence level, intervals with smaller radii are constructed).

2. FIXED RADIUS CONFIDENCE SETS

2.1. Improved Confidence Sets for the Mean

In this section we consider fixed radius confidence sets centered at estimators which shrink toward a linear subspace. For a fairly general class of confidence sets we obtain dominance results similar to those of Hwang and Casella (1984); that is, we prove that these recentered sets have uniformly higher coverage probability than the usual confidence set.

Let X be an observation from a p-variate normal distribution with mean vector θ and covariance matrix $\sigma^2 I$. (Generalization to an arbitrary, known covariance matrix Σ is straightforward, and will be treated later in this section.) Define the estimator $\delta^+(X)$ by

$$\delta^+(X) = [1 - (a\sigma^2/|X|^2)]^+ X \tag{2.1}$$

where a is a positive constant, "+" denotes positive part, and $|\cdot|^2$ is the Euclidean norm. This is the well known positive-part James-Stein estimator, which shrinks the maximum likelihood estimator towards zero.

We consider estimators of the form .

$$\delta^A(X) = AX + \delta^+[(I-A)X] \, , \tag{2.2}$$

where A is a p×p symmetric, idempotent matrix, and confidence sets of the form

$$C_{\delta^A} = \{\theta\colon |\theta-\delta^A(X)| \le c\sigma\} \, . \tag{2.3}$$

The choice of the matrix A is usually based on prior information, reflecting the belief that AX is a reasonable estimator of θ. If $0 \le a \le 2(p-q-2)$, where q is the rank of A, $\delta^A(X)$ is a minimax estimator of θ under squared error loss. Moreover, the region of significant risk improvement is the region where $|(I-A)\theta|/\sigma$ is small. Comparing this to the estimator $\delta^{+}(X)$, which yields significant risk improvement only when $|\theta|/\sigma$ is small, shows that $\delta^{A}(X)$ has a wider region of significant improvement, and is very efficient if the prior belief is true.

The performance of the confidence set C_{δ^A} parallels that of $\delta^{A}(X)$. As will be seen in Theorem 2.1, for suitable choices of a, C_{δ^A} dominates $C^{0}_{X,\sigma}$ in coverage probability. It is also shown that the coverage probability depends on θ only through $|(I-A)\theta|/\sigma$. This, and, for $q = 1$, the numerical results reported in Table 2, show that the region where C_{δ^A} significantly improves upon C^{0} is widened in a way similar to the point estimation case. Consequently, C_{δ^A} is very useful if the prior information is correct, and still dominates $C^{0}_{X,\sigma}$ even if the prior information is incorrect.

As mentioned before, we will also focus on the special case $A = (1/p)\underset{\sim}{1}\underset{\sim}{1}'$, where $\underset{\sim}{1}$ is a $p\times 1$ vector of ones. This choice of A reflects the belief that the θ_i's are close together (or exchangeable), which is the case if the ANOVA null hypothesis is true. The resulting estimator is

$$\delta^{L}(X) = \bar{x}1 + \delta^{+}(X-\bar{x}1) \ , \tag{2.4}$$

the positive part Lindley estimator. The coverage probability of the associated confidence set

$$C_{\delta^L} = \{\theta\colon |\theta-\delta^{L}(X)| \le c\sigma\} \tag{2.5}$$

depends on θ only through $\Sigma_1^p(\theta_i-\bar{\theta})^2/\sigma^2$. Thus, similar to $\delta^{A}(X)$, $\delta^{L}(X)$ significantly widens the region of improvement, obtaining maximal improvement when the θ_i's are close together.

We now proceed to establish the dominance of C_{δ^A} over $C^{0}_{X,\sigma}$ by extending the results of Hwang and Casella (1984), who give conditions for the dominance of $C_{\delta^+} = \{\theta\colon |\theta-\delta^{+}(X)| \le c\sigma\}$ over $C^{0}_{X,\sigma}$. Define the functions $G_q(a,c)$ and $H_q(a,c)$ by

$$G_q(a,c) = \left[\frac{(c/2) + \sqrt{(c/2)^2+a}}{\sqrt{a}}\right]^{p-q-2} e^{-c\sqrt{a}/2} ,$$

$$H_q(a,c) = \left[\frac{c + \sqrt{c^2+a}}{\sqrt{a}}\right]^{p-q-1} e^{-c\sqrt{a}} \left[\frac{\sqrt{c^2+4a} - c}{2\sqrt{a}}\right] . \tag{2.6}$$

Lemma 2.1. (Hwang and Casella, 1984): If $p \geq 3$, the confidence set

$$C_{\delta^+} = \{\theta: |\theta-\delta^+(X)| \leq c\sigma\} \tag{2.7}$$

has higher coverage probability than $C^0_{X,\sigma}$ for every θ provided $a > 0$ and $c > 0$ satisfy $G_0(a,c) \geq 1$ and $H_0(a,c) \geq 1$.

The dominance of C_{δ^A} over $C^0_{X,\sigma}$ can now be established by using a transformation to reduce the problem to that of Lemma 2.1.

Theorem 2.1. Let A be a symmetric, idempotent matrix of rank q, $p-q > 2$. The confidence set C_{δ^A} has higher coverage probability than $C^0_{X,\sigma}$ for every θ provided $0 < a \leq a_0$, where a_0 is the minimum of the two unique solutions to

$$G_q(a,c) = 1 \quad \text{and} \quad H_q(a,c) = 1 . \tag{2.8}$$

Consequently, since C_{δ^A} and $C^0_{X,\sigma}$ have the same volume, C_{δ^A} uniformly dominates $C^0_{X,\sigma}$. Furthermore, the coverage probability of C_{δ^A} depends on θ only through $|(I-A)\theta|/\sigma$.

Remark. Values of a_0 for $q = 1$ and $\alpha = .1$ and $.05$ are given in Table 1. These values of a_0 do not reach the value $p - 3$, which is the optimal choice for the point estimation problem. However, by comparing Tables 2 and 3, it can be seen that the differences in coverage probabilities are minimal.

<u>Proof</u>. Without loss of generality, let $\sigma^2 = 1$ (simply make the transformation $Y = X/\sigma$). Write

$$\begin{aligned} |\theta-\delta^A(X)|^2 &= |[A\theta-AX] + \{(I-A)\theta - \delta^+[(I-A)X]\}|^2 \\ &= |A(\theta-X)|^2 + |(I-A)\theta - \delta^+[(I-A)X]|^2 \quad , \end{aligned} \tag{2.9}$$

where the second equality follows from the fact that A is symmetric idempotent and consequently satisfies $A'(I-A) = 0$. Furthermore, there exists an orthogonal matrix P satisfying $PAP' = D_q$, where D_q is a diagonal matrix whose first q diagonal elements are 1 and last p-q are zero. Define

$$Y = PX \ , \quad \eta = P\theta \ . \tag{2.10}$$

It follows that $Y \sim N(\eta, I)$ and we have

$$|\theta-\delta^A(X)|^2 = \sum_{i=1}^{q} (\eta_i - Y_i)^2 + \sum_{i=q+1}^{p} \{\eta_i - [1-(a/S_Y)]^+ Y_i\}^2 \ , \tag{2.11}$$

where $S_Y = \Sigma_{i=q+1}^{p} Y_i^2$. Using the facts that the Y_i's are independent and $\Sigma_{i=1}^{q} (\eta_i - Y_i)^2 \sim \chi_q^2$, we have

$$P_\theta[|\theta-\delta^A(X)|^2 \le c^2] = \int_0^{c^2} P_\eta\left(\sum_{i=q+1}^{p} |\eta_i - [1-(a/S_Y)]^+ Y_i|^2 \le c^2 - t\right) g_q(t)dt, \tag{2.12}$$

where $g_q(t)$ is the pdf of a χ_q^2 random variable. The theorem will be established if we can show that for every b satisfying $0 < b^2 \le c^2$,

$$P_\eta\left(\sum_{i=q+1}^{p} \{\eta_i - [1-(a/S_Y)]^+ Y_i\}^2 \le b^2\right) > P_\eta\left(\sum_{i=q+1}^{p} (\eta_i - Y_i)^2 \le b^2\right) , \tag{2.13}$$

since substituting the right-hand side of (2.13) for the integrand in (2.12) yields

$$P_\theta(|\theta-\delta^A(X)|^2 \le c^2) > \int_0^{c^2} P_\eta\left(\sum_{i=q+1}^{p} (\eta_i - Y_i)^2 \le c^2 - t\right) g_q(t)d = P_\theta(|\theta - X|^2 \le c^2). \tag{2.14}$$

To establish (2.13) we use Lemma 2.1. Since $(Y_{q+1},\ldots,Y_p)$ is distributed as a (p-q)-variate normal random variable with mean $(\eta_{q+1},\ldots,\eta_p)$ and identity covariance matrix, by Lemma 2.1 it is sufficient to establish that $G_q(a,b) \geq 1$ and $H_q(a,b) \geq 1$ for $0 < a \leq a_0$ and $0 < b^2 \leq c^2$.

We will only give details for the proof that $G_q(a,b) \geq 1$ for $0 < b^2 \leq c^2$, the proof being similar for $H_q(a,c)$. Note that, for each value of b, the function $G_q(a,b)$ is decreasing in a. Hence, it is sufficient to establish that $G_q(a^*,b) \geq 1$, where a^* satisfies $G_q(a^*,c) = 1$. It is straightforward to check that $(\partial/\partial b)\log G_q(a,b)$ is strictly decreasing in b. Hence, $G_q(a^*,b)$ either strictly decreases to zero in b or strictly increases to a unique maximum and then strictly decreases to zero. However, the former case is impossible since $G_q(a^*,0) = G_q(a^*,c) = 1$. Hence, the latter case implies $G_q(a^*,b) > 1$ for $0 < b^2 < c^2$.

Finally, from (2.12), an orthogonal transformation will show that the coverage probability is a function only of $\sum_{i=q+1}^{p}\eta_i^2 = |(I-A)\theta|^2$. ‖

It is straightforward to extend the results of Theorem 2.1 to the case when X has an arbitrary, known covariance matrix Σ. In this case, the usual confidence set is

$$C_\Sigma^0 = \{\theta:\ (\theta-X)'\Sigma^{-1}(\theta-X) \leq c^2\}\ . \tag{2.15}$$

For a given matrix A, define A^* by

$$A^* = \Sigma^{\frac{1}{2}}\ A\Sigma^{-\frac{1}{2}} \tag{2.16}$$

and

$$\delta^{A^*}(X) = A^*X + \left(1 - \frac{a}{X'(I-A^*)'\Sigma^{-1}(I-A^*)X}\right)^+(I-A^*)X\ . \tag{2.17}$$

The confidence set associated with $\delta^{A*}(X)$ is

$$C_{\delta^{A*}} = \{\theta: [\theta-\delta^{A*}(X)]'\Sigma^{-1}[\theta-\delta^{A*}(X)] \le c^2\} . \tag{2.18}$$

The following corollary shows that $C_{\delta^{A*}}$ is a uniform improvement over C_{Σ}^{0}.

Corollary 2.1. The confidence set $C_{\delta^{A*}}$ has uniformly higher coverage probability than C_{Σ}^{0} provided

i) A is symmetric, idempotent of rank q, $p-q>2$, and

ii) $0<a\le a_0$, where a_0 is the unique solution to (2.8).

Proof. The transformation $Y = \Sigma^{-\frac{1}{2}}X$ reduces this to the case of Theorem 2.1. ‖

2.2 Applications

In this section we consider various cases of the estimators and confidence sets constructed in Section 2.1. We pay particular attention to the type of prior information which may be useful in achieving the greatest possible improvement.

Example 1. The Unbalanced One-Way Analysis of Variance

In a one-way ANOVA model, it is assumed that there are p treatments characterized by the levels $\theta_1,\cdots,\theta_p$, and there are n_i i.i.d. $N(\theta_i,\sigma^2)$ observations $X_{i1},\cdots,X_{in_i}$. The variance, σ^2, is assumed to be known.

By sufficiency, we can consider only procedures depending on $\bar{X}_i = \frac{1}{n_i}\Sigma_{i=1}^{N_i}X_{ij}$, $1\le i\le p$. The vector $\bar{X} = (\bar{X}_1,\cdots,\bar{X}_p)'$ has a multivariate normal distribution with mean $(\theta_1,\cdots,\theta_p)'$ and covariance matrix $\Sigma = \sigma^2 D_n^{-1}$, $D_n = \mathrm{diag}(n_1,\cdots,n_p)$. The usual confidence set is

$$C_{\bar{X},\sigma}^{0} = \{\theta: (\theta-\bar{X})'D_n(\theta-\bar{X}) \le c^2\sigma^2\} . \tag{2.19}$$

Corollary 2.1 provides better confidence sets, however.

Consider the situation when prior information indicates θ_i's are very likely to be close. Under the assumption that $\theta_i = \theta_j$, $\forall i, j$, the classical estimator (the uniformly minimum variance unbiased estimator and the maximum likelihood estimator) of the common value θ is $\bar{\bar{x}} = \frac{1}{N}\Sigma_{i=1}^{p}\Sigma_{i=1}^{n_i}X_{ij}$, where $N = \Sigma n_i$. In this situation, the reasonable estimator to use is one which shrinks $\bar{X}$ to an estimate of the common θ. If we take $A = D_n^{\frac{1}{2}}\underset{\sim}{1}\underset{\sim}{1}'D_n^{\frac{1}{2}}/N$, we have $A^* = \underset{\sim}{1}\underset{\sim}{1}'D_n/N$, and

$$\delta^{A^*}(\bar{X}) = \bar{\bar{x}}\underset{\sim}{1} + \left[1 - \frac{a\sigma^2}{\sum_{i=1}^{p} n_i(\bar{X}_i - \bar{\bar{x}})^2}\right]^+ (\bar{X} - \bar{\bar{x}}\underset{\sim}{1}) \quad , \tag{2.20}$$

a version of the positive-part Lindley estimator. Note that A is a symmetric, idempotent matrix, and hence Corollary 2.1 shows that the confidence set

$$C_{\delta^{A^*}} = \{\theta: [\theta - \delta^{A^*}(\bar{X})]'D_n[\theta - \delta^{A^*}(\bar{X})] \le c^2\sigma^2\} \tag{2.21}$$

uniformly dominates $C^0_{\bar{X},\sigma}$ provided $0 \le a \le a_0$, where a_0 is the largest value which satisfies $G_1(a,c) \ge 1$ and $H_1(a,c) \ge 1$. For $a = .1$ and $.05$, the values of a_0 are given in Table 1.

Another, somewhat more vague, prior specification which may be useful is to suppose that θ is a multiple of a known vector, i.e., $\theta = \alpha r$, where α is an unknown scalar and r is a known vector. If $r = \underset{\sim}{1}$, this case reduces to the previous situation. Under the model $\theta = \alpha r$, the classical estimator (i.e., the uniformly minimum unbiased estimator and the maximum likelihood estimator) of θ is

$$\delta^r(\bar{X}) = rr'\Sigma^{-1}\bar{X}/r'\Sigma^{-1}r \tag{2.22}$$

with i^{th} component

$$\delta_i^r(\bar{X}) = r_i \sum_j r_j n_j \bar{X}_j / \sum_j n_j r_j^2 \quad .$$

It is therefore reasonable to consider δ^{A*} that shrinks $\bar{X}$ toward $\delta^{r}(\bar{X})$. If we choose

$$A = D_n^{\frac{1}{2}} rr' D_n^{\frac{1}{2}} / r' D_n r \quad , \tag{2.23}$$

which is symmetric and idempotent, we have $A^* = D_n^{-\frac{1}{2}} A D_n^{\frac{1}{2}} = rr'D_n/r'D_n r$. Corollary 2.1 then shows that a confidence set of the form (2.21) centered at

$$\delta^{A*}(X) = \delta^{r}(\bar{X}) + \left[1 - \frac{a\sigma^2}{\sum_{i=1}^{p} n_i[\bar{X}_i - \delta_i^{r}(\bar{X})]^2}\right]^{+} [\bar{X} - \delta^{r}(\bar{X})]$$

dominates one centered at $\delta^{r}(\bar{X})$.

Example 2. θ Lies in a Linear Subspace

We now consider a more general form of restriction, one in which the prior information indicates that θ lies in a linear subspace of the parameter space. This example includes the case of the general linear model (with known variance) as a special case, if X, below, is taken to be the least squares estimator.

Assume $X \sim N(\theta, \Sigma)$ where Σ is a nonsingular known matrix. Suppose that the prior information indicates that θ lies in the plane

$$L_H = \{\theta : H\theta = 0\} \quad ,$$

where H is a $k \times p$ matrix with rank k. Assuming $\theta \epsilon L_H$, it can be shown that the classical estimator (the uniformly minimum variance unbiased estimator and the maximum likelihood estimator) of θ is A^*X with

$$A^* = I - \Sigma H'(H\Sigma H')^{-1} H \quad .$$

In Corollary 2.1, let $A = I - \Sigma^{-\frac{1}{2}} H'(H\Sigma H')^{-1} H \Sigma^{\frac{1}{2}}$, which is symmetric and idempotent. The resulting estimator is

$$\delta^{A*}(X) = A^*X + \left[1 - \frac{a}{X'H'(H\Sigma H')^{-1}HX}\right]^{+} \Sigma H'(H\Sigma H')^{-1} HX \quad ,$$

and it follows that the confidence set centered at this estimator uniformly dominates the one centered at A^*X provided $0<a\le a_0$, where a_0 satisfies $G_k(a,c)\ge 1$ and $H_k(a,c)\ge 1$.

If, instead, prior belief indicates that θ is near $L_{H^*} = \{\theta\colon H\theta=m\}$, which is assumed to be a nonempty set with H being as above and m a known k-component vector, we can proceed as follows. Assume that $\theta_0 \epsilon L_{H^*}$, i.e., $H\theta_0 = m$. By transforming $X' = X-\theta_0$ and $\theta' = \theta-\theta_0$, the problem is reduced to the above setting. It then follows that an improved confidence set can be constructed by centering at the estimator

$$\delta^{A^*}(X)=[X-\Sigma H'(H\Sigma H')^{-1}(HX-m)] + \left[1 - \frac{a}{(HX-m)'(H\Sigma H)^{-1}(HX-m)}\right]^+ \Sigma H'(H\Sigma H')^{-1}(HX-m) \ .$$

3. VARIABLE RADIUS CONFIDENCE SETS

The confidence sets considered in Section 2 are all of fixed radius and, hence, afford no volume reduction over the usual set. Although sets such as C_{δ^L} of (2.5) yield uniformly higher coverage probability, an experimenter must report the same confidence coefficient as reported if the usual set, $C^0_{X,\sigma}$, had been used. Thus, to the experimenter, there is no tangible evidence that C_{δ^L} should be preferred over $C^0_{X,\sigma}$.

Since the coverage probability of C_{δ^L} can be much greater than that of $C^0_{X,\sigma}$, there seems to be room for decreasing its volume without giving up dominance in coverage probability. The construction of such confidence sets is the focus of this section.

We confine our attention to the estimator

$$\delta^L(X) = \bar{x}\underset{\sim}{1} + \left(1 - \frac{a\sigma^2}{\Sigma(X_i-\bar{x})^2}\right)^+ (X-\bar{x}\underset{\sim}{1}) \quad , \tag{3.1}$$

the positive part Lindley estimator. Through the use of transformations like those in Section 2, the results of this section can be generalized to include estimators such as δ^A of (2.2), and associated confidence sets. However, we will not consider such generalizations here.

In Section 3.1 we derive, through the use of a modified empirical Bayes argument, a variable radius confidence set based on $\delta^L(X)$. It is shown that this set has uniformly smaller volume than the usual confidence set and, in fact, can provide significant volume reduction. Also, the exact formula for the coverage probability is derived. Although dominance in coverage probability is not demonstrated analytically, strong numerical evidence is presented which shows that the empirical Bayes confidence set is superior to the usual set. Section 3.2 deals with the case of unknown σ^2, where the empirical Bayes sets are modified by replacing σ^2 by an estimate. Again, these sets yield a reduction in volume, and (based on numerical evidence) also dominate the usual confidence set in coverage probability.

3.1. Empirical Bayes Confidence Sets

We now consider confidence sets of the form

$$\{\theta: |\theta-\delta^{L}(X)|^2 \le v(X,\sigma)\} \tag{3.2}$$

where v is a nondecreasing function of $|X-\bar{x}\underset{\sim}{1}|$. It is a difficult task to find a function v that will yield a confidence set dominating the usual one in coverage probability. To get some idea of what form such a function will take, we use an empirical Bayes argument. We begin by deriving a Bayes rule against the loss function

$$L_k(\theta,C) = k \text{ Volume } (C) - I_C(\theta) \quad , \tag{3.3}$$

where $I_C(\theta) = 1$ if $\theta\varepsilon C$ and zero otherwise. For this loss function, the usual confidence set

$$C^{0}_{X,\sigma} = \{\theta: |\theta-X| \le c\sigma\}$$

is minimax if $k=k_0=\exp(-c^2/2\sigma^2)/2\pi\sigma^2)^{p/2}$. Since k_0 is the only value of k for which $C^0_{X,\sigma}$ is minimax, it seems reasonable to use this value in deriving a Bayes rule. Furthermore, Casella and Hwang (1982, Theorems 2.1, 2.2, and 2.3) argue that a rule that is minimax with respect to L_k is likely to dominate $C^0_{X,\sigma}$.

An empirical Bayes argument has been previously used (Casella and Hwang, 1983) to derive confidence sets centered at the positive-part James-Stein estimator. There, normal priors were used in a way parallel to the derivation of the James-Stein point estimator in Efron and Morris (1973). More recently, Morris (1983) has used empirical Bayes arguments to construct improved confidence intervals. Out goal here is to derive confidence sets centered at $\delta^L(X)$ so, naturally, we consider Lindley's two-stage prior, namely

$$X|\theta,\sigma^2 \sim N(\theta,\sigma^2 I)$$
$$\theta|\mu,\tau^2 \sim N(\mu\underset{\sim}{1},\tau^2 I) \tag{3.4}$$
$$\mu \sim \text{Uniform}(-\infty,\infty) \quad .$$

The improper prior on the scalar μ can be interpreted as an approximation to an $n(0,\lambda^2)$ density, where λ^2 is much larger than σ^2 or τ^2.

From Joshi (1969) or Faith (1976), it follows that the Bayes rule against L_k is

$$C^B = \{\theta:\pi(\theta|X) \geq k\} \ , \tag{3.5}$$

where $\pi(\theta|X)$ is the posterior density of θ. Direct calculation shows that the posterior distribution is $N[\delta^B(X),V^{-1}]$, where

$$\delta^B(X) = \bar{x}\underset{\sim}{1} + \frac{\tau^2}{\sigma^2+\tau^2}(X-\bar{x}\underset{\sim}{1})$$
$$V = \left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right) I - \left(\frac{1}{p\tau^2}\right) \underset{\sim}{1}\underset{\sim}{1}' \ . \tag{3.6}$$

Setting $k=k_0$, the Bayes rule against L_{k_0} can be written (after some algebra) as

$$C^B = \{\theta: [\theta-\delta^B(X)]'V[\theta-\delta^B(X)] \leq v_B(\tau^2,\sigma^2)\} \tag{3.7}$$

where

$$v_B(\tau^2,\sigma^2) = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\{c^2 - (p-1)\log[\tau^2/(\tau^2+\sigma^2)]\} \quad . \tag{3.8}$$

The prior variance, τ^2, is usually unknown, and the empirical Bayesian will replace τ^2 by an estimate taken from the marginal density of X. Marginally, $\Sigma_{i=1}^p(X_i-\bar{x})^2 \sim (\sigma^2+\tau^2)\chi^2_{p-1}$, and it follows that

$$E\left\{1 - \frac{(p-3)\sigma^2}{\Sigma(X_i-\bar{x})^2}\right\} = \frac{\tau^2}{\tau^2+\sigma^2} \quad . \tag{3.9}$$

The empirical Bayes strategy is to replace $\tau^2/(\tau^2+\sigma^2)$ in v^B by its unbiased estimate. However, the unbiased estimate may be negative, which is undesirable. Hence, we truncate the unbiased estimate, and replace v^B by the modified empirical Bayes estimate

$$v^E(X,\sigma) = \sigma^2 T[c^2 - (p-1)\log T] \tag{3.10}$$

where

$$T = \begin{cases} 1 - \dfrac{(p-3)\sigma^2}{\Sigma(X_i-\bar{x})^2} & \text{if} \quad \dfrac{\Sigma(X_i-\bar{x})^2}{\sigma^2} \geq c^2 \\ 1 - \dfrac{p-3}{c^2} & \text{if} \quad \dfrac{\Sigma(X_i-\bar{x})^2}{\sigma^2} < c^2 \end{cases} . \tag{3.11}$$

We assume that $c^2 > p-3$ so that log T is defined. This an extremely minor assumption since $c^2 = p-3$ would give $C^0_{X,\sigma}$ a confidence coefficient of approximately .3 .

Now consider the left-hand side of the inequality in (3.7). Direct calculation shows

$$[\theta-\delta^B(X)]'V[\theta-\delta^B(X)] = |\theta-\delta^B(X)|^2 - \frac{p\sigma^2}{\tau^2+\sigma^2}(\bar{x}-\bar{\theta})^2 \quad . \tag{3.12}$$

If we use the empirical Bayes strategy on (3.12), this will lead us to a very complicated confidence set. It not only will be hard to interpret (it may not be convex), but also its coverage probability will be quite hard to evaluate. Thus, we take a simpler alternative, and merely drop the last term in (3.12) (which also decreases the volume of the confidence set). Upon replacing $\delta^B(X)$ by its estimate, $\delta^L(X)$, we obtain our recommended confidence set

$$C^E_{X,\sigma} = \{\theta: |\theta-\delta^L(X)|^2 \leq v^E(X,\sigma)\} \quad . \tag{3.13}$$

The coverage probability of $C^E_{X,\sigma}$ can be evaluated by using a decomposition similar to that used in Section 2, Theorem 2.1. Taking $\sigma^2 = 1$ we have

$$P_\theta(\theta\varepsilon C^E_{X,\sigma}) = \int_0^{c^2} P_\eta\{|\eta-\delta^+(Y)|^2 \le [v^*(|Y|)-t]^+\}g_1(t)dt \tag{3.14}$$

where $Y\sim N_{p-1}(\eta,I)$, $\eta = \theta-\bar{\theta}\underset{\sim}{1}$, $\delta^+(Y) = [1-(p-3)/|Y|^2]^+Y$, $g_1(t)$ is the density of a χ^2_1 random variable, and

$$v^*(|Y|) = \left(1 - \frac{p-3}{\max\{|Y|^2,c^2\}}\right)\left[c^2 - (p-1)\log\left(1 - \frac{p-3}{\max\{|Y|^2,c^2\}}\right)\right] .$$

To evaluate the integrand of (3.14), transform to the spherical coordinates $r = |Y|$, $\cos\beta = \eta'Y/|\eta||Y|$. We then have

$$\{|\eta-\delta^+(Y)|^2 \le [v^*(|Y|)-t]^+\} = \{r^2\gamma^2(r) - 2r\gamma(r)|\eta|\cos\beta + |\eta|^2 \le [v^*(r)-t]^+\}, \tag{3.15}$$

where $\gamma(r) = [1-(a/r^2)]^+$.

Some algebra will show that this last set can be written as

$$\{(r,\beta):\ r\varepsilon S_{\eta,t},\ \cos\beta > h(r,t)\} \quad , \tag{3.16}$$

where $S_{\eta,t}$ is either an interval, or the union of two disjoint intervals, and is defined by

$$S_{\eta,t} = \left\{r:\ \{|\eta|-[r-(a/r)]^+\}^2 \le [v^*(r)-t]^+\right\} \quad , \tag{3.17}$$

and

$$h(r,t) = \begin{cases} \max\left\{\dfrac{r^2\gamma^2(r)+|\eta|^2-[v^*(r)-t]^+}{2r|\eta|\gamma(r)} , -1\right\} & \text{if } r\gamma(r) \ne 0 \\ -1 & \text{if } r\gamma(r) = 0 \end{cases} . \tag{3.18}$$

We thus have the following representation for the coverage probability of $C^E_{X,\sigma}$

<u>Theorem 3.1.</u> If $p \ge 3$, for $|\theta-\bar{\theta}\underset{\sim}{1}| > 0$,

$$P_\theta(\theta\varepsilon C^E_{X,\sigma}) = K\int_0^{c^2}\int_{S_{\eta,t}}\int_0^{\cos^{-1}(h)} r^{p-2}(\sin\beta)^{p-3}$$

$$\times \exp\{-(r^2-2r|\eta|\cos\beta+|\eta|^2)/2\}g_1(t)d\beta drdt \quad , \tag{3.19}$$

where $K^{-1} = \sqrt{\pi}\Gamma[(p-2)/2]2^{(p-3)/2}$, $\eta = |\theta-\bar{\theta}\underset{\sim}{1}|/\sigma$, and $g_1(t)$ is the density of a χ_1^2 random variable. If $|\theta-\bar{\theta}\underset{\sim}{1}|=0$, then $P_\theta(\theta\epsilon C^E_{X,\sigma}) = \int_0^{c^2} P[\chi^2_{p-1} \le r_+(0,t)]g_1(t)dt$, where $r_+(\eta,t) = \max\{r: r\epsilon S_{\eta,t}\}$.

<u>Proof</u>. If $|\theta-\bar{\theta}\underset{\sim}{1}|=0$, the set $\{Y: |\eta-\delta^+(Y)|^2 \le [v^*(|Y|)-t]^+\}$ clearly contains $Y=0$. Hence $S_{0,t}$ is an interval and the result follows. The result for $|\theta-\bar{\theta}\underset{\sim}{1}| > 0$ is easily established by carrying out the spherical transformation. ||

If v^E is constant, as in Section 2, then it is possible to express (3.17) as an interval. The fact that this cannot be done when v^E is nonconstant is the major reason why dominance of $C^E_{X,\sigma}$ over $C^0_{X,\sigma}$ in coverage probability cannot be established analytically. Formula (3.19) has been evaluated numerically, however. For $a=p-3$ and $\alpha=.1$, coverage probabilities of $C^E_{X,\sigma}$ are presented in Table 4. We choose to use $a=p-3$, rather than $a=a_0$, because this value produces a better point estimator than $a=a_0$, and is more readily available to an experimenter. The numerical evidence shows that, with the exception of $p=4$ and $\alpha=.1$, $C^E_{X,\sigma}$ provides a uniform improvement over $C^0_{X,\sigma}$. Moreover, the failure of $C^E_{X,\sigma}$ is so slight (for example, a minimum coverage probability of .891 for $a=.1$) that for all intents and purposes, $C^E_{X,\sigma}$ can be regarded as a $1-\alpha$ confidence set.

The important feature of $C^E_{X,\sigma}$, however, is that it provides a reduction in volume over $C^0_{X,\sigma}$. This is established in the following theorem.

<u>Theorem 3.2.</u> If $c^2>p-1$, the radius of $C^E_{X,\sigma}$ is uniformly (in $|X-\bar{x}\underset{\sim}{1}|$) smaller than that of $C^0_{X,\sigma}$. Hence, $C^E_{X,\sigma}$ has uniformly smaller volume.

<u>Proof</u>. From (3.10),

$$v^E(X,\sigma) = \sigma^2 T[c^2-(p-1)\log T] ,$$

where T is defined in (3.11). Differentiation shows $(\partial v^E/\partial T) \geq 0$, which implies that v^E is a nondecreasing function of T. Since T is a nondecreasing function of $|X-\bar{x}\underset{\sim}{1}|$, and $0 < T \leq 1$, it follows that v^E is nondecreasing in $|X-\bar{x}\underset{\sim}{1}|$, and is bounded by $c^2\sigma^2$. ||

To get an idea of the amount of possible improvement, Table 5 gives values of the ratio of the radii of $C^E_{X,\sigma}$ to $C^0_{X,\sigma}$, i.e.,

$$\frac{\text{Radius of } C^E_{X,\sigma}}{\text{Radius of } C^0_{X,\sigma}} = \left[\frac{T[c^2-(p-1)\log T]}{c^2}\right]^{\frac{1}{2}} . \tag{3.20}$$

As can be seen in Table 5, the amount of possible improvement can be substantial if $|X-\bar{x}\underset{\sim}{1}|$ is small. We also note that, in terms of volume reduction (rather than radius reduction) the improvement is even greater. The ratio of volumes is obtained by raising (3.20) to the p^{th} power and, hence, is smaller than the ratio of the radii.

3.2. The Case of Unknown Variance

We assume now that σ^2 is unknown, but an estimate s^2 of σ^2, independent of X, is available, where $s^2 \sim (\sigma^2/\nu)\chi^2_\nu$. The usual $1-\alpha$ confidence set for θ is

$$C^0_{X,s} = \{\theta: |\theta-X|^2 \leq s^2c^2\} , \tag{3.21}$$

where c^2 satisfies $P(F_{p,\nu} \leq c^2/p) = 1-\alpha$, where $F_{p,\nu}$ denotes an F-random variable with p and ν degrees of freedom.

In order to obtain an improved confidence set, it should be possible to proceed as in Section 3.1, and consider a modified empirical Bayes set.

However, it is no longer clear as to which priors would lead us to reasonable empirical Bayes sets. An obvious choice would be to use a conjugate prior, but such calculations lead to enormously complicated sets with no easily discernible optimality properties. One disconcerting fact is the following: consider the simple known variance case $X|\theta \sim N(\theta,\sigma^2 I)$ and $\theta|\tau^2 \sim N(0,\tau^2 I)$. As $\tau^2 \to \infty$, the Bayes set converges to $C^0_{X,\sigma}$ and the Bayes risk converges to the risk of $C^0_{X,\sigma}$. This no longer occurs in the unknown variance case. For the conjugate priors $(\theta|\sigma^2,\tau^2 \sim N(0,\sigma^2,\tau^2 I), \sigma^2 \sim$ Inverse Gamma) as $\tau^2 \to \infty$, neither the Bayes set nor the Bayes risk converges to that of $C^0_{X,s}$.

Thus, we consider the simpler alternative of replacing σ^2 in $C^E_{X,\sigma}$ by its estimate, s^2, and form the confidence set

$$C^E_{X,s} = \{\theta\colon\ |\theta-\delta^L(X,s)|^2 \le v^E(X,s)\} \tag{3.22}$$

where

$$v^E(X,s) = s^2 T_\nu[c^2-(p-1)\log T_\nu] \quad ,$$

$$T_\nu = \begin{cases} 1 - \dfrac{a_\nu s^2}{\Sigma(X_i-\bar{x})^2} & \text{if} \quad \dfrac{\Sigma(X_i-\bar{x})^2}{s^2} \ge c^2 \quad , \\[2ex] 1 - \dfrac{a_\nu}{c^2} & \text{if} \quad \dfrac{\Sigma(X_i-\bar{x})^2}{s^2} < c^2 \quad , \end{cases} \tag{3.23}$$

$$\delta^L(X,s) = \bar{x}\underset{\sim}{1} + \left(1 - \frac{a_\nu s^2}{\Sigma(X_i-\bar{x})^2}\right)^+ (X-\bar{x}\underset{\sim}{1}) \quad ,$$

and

$$a_\nu = \nu(p-3)/(\nu+2) \quad .$$

The choice $a=a_\nu$ again reflects the fact that a_ν results in an optimal point estimator. Also, if we assume the structure of (3.4) and merely regard σ^2 as a nuisance parameter, we have (similar to (3.9))

$$\lim_{\nu\to\infty} E\left[1 - \frac{a_\nu s^2}{\Sigma(X_i-\bar{x})^2}\right] = \frac{\tau^2}{\tau^2+\sigma^2} \quad . \tag{3.24}$$

Thus, $C^E_{X,s}$ retains somewhat of an empirical Bayes flavor.

Again, dominance of coverage probability of $C^E_{X,s}$ over $C^0_{X,s}$ could not be obtained analytically. The major reason for this is the same as in the known variance case: the limits of integration could not be solved for explicitly. However, the exact formula for the coverage probability of $C^E_{X,s}$ is straightforward to derive, and can be easily obtained for the formula for the coverage probability of $C^E_{X,\sigma}$.

<u>Theorem 3.3.</u> The coverage probability of $C^E_{X,s}$ is

$$P_\theta(\theta\epsilon C^E_{X,s}) = \int_0^\infty P[|\theta-\delta^L(X,t)|^2 \le v^E(X,t)]g_\nu(t)dt \ , \tag{3.25}$$

where $g_\nu(t)$is the density of a χ^2_ν random variable, and δ^L and v^E are given in (3.23).

<u>Proof</u>. The formula follows immediately from the independence of X and s^2. ||

The integrand in equation (3.25) can be evaluated using Theorem 3.1. Values of the coverage probability have been evaluated numerically for $p=4,10$, $\nu=2,5,10,20,30$ and $\alpha=.1$, and are presented in Table 6. With the exception of $p=4$, and a few other cases where $\nu=2$, $C^E_{X,s}$ demonstrates almost uniform dominance in coverage probability over $C^0_{X,s}$. Again, when $C^E_{X,s}$ fails, the failure is so slight that it is reasonable to treat $C^E_{X,s}$ as a $1-\alpha$ confidence set.

$C^E_{X,s}$ does provide a uniform reduction in volume over $C^0_{X,s}$ and, similar to Theorem 3.2, we have the following theorem.

Theorem 3.4. If $c^2 > p-1$, then Volume $(C^E_{X,s})$ < Volume $(C^0_{X,s})$ for all values of X and s^2.

Proof. From (3.23), it is clear that the ratio of the radii of $C^E_{X,s}$ to $C^0_{X,s}$ (and hence the ratio of the volumes) is a function of the data only through $\Sigma(X_i-\bar{x})^2/s^2$. The theorem is then established in a manner similar to that of Theorem 3.2. ||

Selected values of the radius ratio are presented in Table 7. While the reduction in volume is not as much as for the known variance case, $C^E_{X,s}$ can provide good volume reduction if $\Sigma(X_i-\bar{x})^2/s^2$ is small.

4. ESTIMATION OF CONTRASTS IN THE ONE-WAY ANALYSIS OF VARIANCE

Often, in the analysis of variance, an experimenter is not only interested in testing the overall hypothesis that the treatment means are equal, but also in testing or estimating linear combinations of the means. The confidence sets developed in the previous two sections can be readily adapted to such situations. In fact, the variable radius confidence sets of Section 3 provide a procedure that is a uniform improvement over the S-method of Scheffé (1959).

Recall the setup of the one-way analysis of variance, mentioned in Section 2, Example 1. For any $p \times 1$ vector a, we call $\psi_\theta(a) = \Sigma a_i\theta_i$ a *comparison* of the means $\theta_1,\cdots,\theta_p$. In addition, if $\Sigma a_i = 0$, $\psi_\theta(a)$ is called a *contrast*. The classic estimator of $\psi_\theta(a)$ is $\psi_X(a) = \Sigma a_i\bar{X}_i$. The multiple comparison procedure of Scheffé (1959), the S-method, can be summarized as follows.

Theorem 4.1. (Scheffé):

a) The probability is $1 - \alpha$ that simultaneously

$$|\psi_X(a) - \psi_\theta(a)| \le \left(pF_{\alpha,p,\nu}s^2\sum_i a_i^2/n_i\right)^{\frac{1}{2}} \tag{4.1}$$

for all vectors a.

b) The probability is $1 - \alpha$ that simultaneously

$$|\psi_X(a) - \psi_\theta(a)| \le \left((p-1)F_{\alpha,p-1,\nu}s^2\Sigma a_i^2/n_i\right)^{\frac{1}{2}} \tag{4.2}$$

for all vectors a such that $\Sigma a_i = 0$.

The Scheffé intervals can be used either for estimation or testing. They are very flexible in their ability to handle all contrasts, and to provide exact probability statements even for unbalanced data. The major drawback of the Scheffé intervals is that they are very conservative. For example, if one is only interested in pairwise differences, it would be better to use Tukey's procedure (sometimes referred to as the Q-method), which provides shorter intervals for the same α level (see also Krishnaiah, 1965).

The Scheffé intervals are based on a confidence ellipse centered at the observations. Since the confidence sets detailed in Section 2 and 3 are improvements over this ellipse, we expect intervals based on these sets to improve upon the Scheffé intervals.

For an estimator $\delta(\bar{X},s)$ of θ, define $\psi_\delta(a)$ by

$$\psi_\delta(a) = \Sigma a_i \delta_i(\bar{X},s) \quad . \tag{4.3}$$

We have the following theorem, which shows the relationship between the improved confidence sets and improved intervals.

<u>Theorem 4.2.</u> Let $D_n = \mathrm{diag}(n_1,\cdots,n_p)$. If the confidence set

$$C_\delta = \left\{\theta: \left(\theta - \delta(\bar{X},s)\right)' D_n \left(\theta - \delta(\bar{X},s)\right) \le v_\alpha^E(\bar{X},s)\right\} \tag{4.4}$$

is a $1-\alpha$ confidence set, then the probability is $1-\alpha$ that simultaneously

$$|\psi_\delta(a) - \psi_\theta(a)| \le\!< [v_\alpha^E(\bar{X},s)\Sigma a_i^2/n_i]^{\frac{1}{2}} \tag{4.5}$$

for all vectors a.

<u>Proof.</u> $P\{|\psi_\delta(a) - \psi_\theta(a)|^2 \le v_\alpha^E(\bar{X},s)\Sigma a_i^2/n_i \text{ for all } a\}$

$$= P\left\{\max_a \frac{a'[\theta-\delta(\bar{X},s)][\theta-\delta(\bar{X},s]'a}{a'D_n^{-1}a} \le v_\alpha^E(\bar{X},s)\right\}$$

$$= P\{[\theta-\delta(\bar{X},s)]'D_n\,[\theta-\delta(\bar{X},s)] \le v_\alpha^E(\bar{X},s)\} \quad . \qquad \|$$

Generalization of this theorem, and the rest of the results in this section, to the case when the covariance matrix is not diagonal is straightforward and will not be dealt with explicitly.

We now will use Theorem 4.2 to construct intervals based on the confidence set $C^E_{X,s}$ of Section 3. For the analysis of variance setup of this section, the appropriate version of $C^E_{X,s}$ is

$$C^E_{\bar{X},s} = \left\{\theta: \left(\theta - \delta^L(\bar{X},s)\right)' D_n\left(\theta - \delta^L(\bar{X},s)\right) \le v^E_\alpha(\bar{X},s)\right\} \quad , \tag{4.6}$$

where $\delta^L(\bar{X},s)$ and $v^E_\alpha(\bar{X},s)$ are defined as in (3.23) with the exceptions that $\Sigma(X_i-\bar{x})^2$ is replaced by $\Sigma n_i(\bar{X}_i-\bar{\bar{x}})^2$ and c^2 is replaced by $pF_{\alpha,p,\nu}$. If $C^E_{X,s}$ is $1-\alpha$ confidence set, then the probability is $1-\alpha$ that simultaneously

$$|\psi_\theta(a) - \psi_{\delta^L}(a)| \le [v^E_\alpha(\bar{X},s)\Sigma a_i^2/n_i]^{\frac{1}{2}} \tag{4.7}$$

for all vectors a. Theorem 3.4 guarantees that if $pF_{\alpha,p,\nu} > p-1$, then $v^E(\bar{X},s) \le s^2 pF_{\alpha,p,\nu}$. Hence, by comparing (4.1) to (4.7), it follows that the intervals of (4.7) are uniformly shorter than the Scheffé intervals. Of course, it was not demonstrated analytically that $C^E_{\bar{X},s}$ is a $1-\alpha$ confidence set. However, the numerical evidence for this case is quite strong (with the exceptions noted in Section 3).

To get an idea of the possible improvement that the intervals in (4.7) can provide over the corresponding S-intervals given in (4.1), Table 7 can be used. For a given p, the entries in Table 7 give the ratio of the lengths of the intervals in (4.7) to the corresponding Scheffé intervals in (4.1).

For the case of simultaneous estimation of contrasts, the argument is similar, the only difference being that the intervals are taken from a p-1 dimensional ellipse. (The space of all contrasts is a p-1 dimensional subspace of the p-dimensional parameter space.) Similar to Theorem 4.2, we have the following theorem for contrasts, which we state without proof.

Theorem 4.3. Let $\bar{X} \sim N(\theta, D_n^{-1})$. Define the $p \times (p-1)$ matrix Q by $Q = (I_{p-1}, -\underset{\sim}{1})'$, and let $Y = Q'X$, $\eta = Q'\theta$. If the confidence set

$$C_Y = \{\eta: [\eta - Q'\delta(\bar{X},s)]'(Q'D_n^{-1}Q)^{-1}[\eta - Q'\delta(\bar{X},s)] \le v_\alpha^E(\bar{X},s)\} \qquad (4.8)$$

is a $1-\alpha$ confidence set, then the probability is $1-\alpha$ that simultaneously

$$|\psi_\delta(a) - \psi_\theta(a)| \le [v_\alpha^E(\bar{X},s)\Sigma a_i^2/n_i]^{\frac{1}{2}} \qquad (4.9)$$

for all vectors a such that $\Sigma a_i = 0$.

For this case our recommended confidence set, and hence simultaneous intervals for all contrasts can again be described by (3.23) with two exceptions. We again replace $\Sigma(X_i - \bar{x})^2$ with $\Sigma n_i(\bar{X}_i - \bar{\bar{x}})^2$, but now we replace c^2 by $(p-1)\, F_{\alpha,p-1,\nu}$.

To see more clearly how we arrive at this new modification v_α^E for the case of contrasts, we can calculate explicitly the ellipse (4.8) of Theorem 4.3 when we use the estimator $\delta^L(\bar{X},s)$. It is easy to check that

$$Q'\delta^L(\bar{X},s) = \left(\frac{a_\nu s^2}{Y'(Q'D_n^{-1}Q)^{-1}Y}\right)^+ Y \overset{\text{def'n}}{=} \delta(Y) \quad ,$$

since $Q'\bar{\bar{x}}\underset{\sim}{1} = 0$ and $\Sigma n_i(\bar{X}_i - \bar{\bar{x}})^2 = Y'(Q'D_n^{-1}Q)^{-1}Y$. Furthermore, $Y = Q'X$ is distributed as a $(p-1)$-dimensional normal random variable with mean $\eta = Q'\theta$ and covariance matrix $\Sigma_Y = \sigma^2(Q'D_n^{-1}Q)^{-1}$. Thus, the intervals in (4.9) will have simultaneous coverage probability at least $1-\alpha$ if the $(p-1)$-dimensional confidence set

$$C_Y = \{\eta: [\eta - \delta(Y)]'\Sigma_Y[\eta - \delta(Y)] \le v_\alpha^E(Y,s)\} \qquad (4.10)$$

is a $1-\alpha$ confidence set.

Note that the estimator $\delta(Y)$ shrinks Y toward zero rather than toward a linear subspace, and thus the associated confidence set is a special case of the sets in Sections 2 and 3, obtained by taking the matrix A=0. Also, if σ^2 is known, then the coverage probability of C_Y can be obtained from Theorem 3.1 by merely setting t=0 in (3.17) and (3.18), and eliminating the integral over t in (3.19). For the case of unknown σ^2, one proceeds as in Theorem 3.3, and integrates this probability against the χ^2_ν density. It is also easy to verify that the coverage probability of C_Y depends on the unknown parameters only through $\eta'\Sigma_Y^{-1}\eta = \theta'Q(Q'D_n^{-1}Q)^{-1}Q'\theta/\sigma^2 = \Sigma n_i(\theta_i-\bar{\theta})^2/\sigma^2$. (Sets such as C_Y, which are centered at estimators that shrink toward a known point rather than a linear subspace, are treated in more detail in Casella and Hwang (1983).)

The coverage probability of C_Y, and hence of the associated intervals, has been evaluated numerically for p=4,10, ν=2,3,5,10,20,30, and a=.1, and are presented in Table 8. With the exception of a few cases when p=4 or ν=2, the coverage probability of C_Y is above .90. Again, the few failures are so slight that it is reasonable to consider C_Y a 1-α procedure.

If we apply Theorem 3.4 to the intervals associated with C_Y, we find that a sufficient condition for these intervals to be uniformly shorter than the corresponding Scheffé intervals is $F_{\alpha,p-1,\nu}>1$. The ratio of the radii of these intervals is given by

$$\left[\frac{v_\alpha^E(\bar{X},s)\Sigma a_i^2/n_i}{(p-1)s^2F_{\alpha,p-1,\nu}\Sigma a_i^2/n_i}\right]^{\frac{1}{2}} = \left[T_\nu\left(1-\frac{\log T_\nu}{F_{\alpha,p-1,\nu}}\right)\right]^{\frac{1}{2}}, \qquad (4.11)$$

where T_ν is described following Theorem 4.3. Values of this ratio for p=4,10, v=2,5,10,20,30, and α=.10 have been calculated, and are given in Table 9. The reduction in length is quite respectable, and is even greater than for the case of estimation of contrasts (compare Tables 7 and 9).

The results of this section also apply to the case of known variance. The relevant intervals are obtained by letting $\nu\to\infty$, in the expressions given in this section. (As $\nu\to\infty$, $s^2\to\sigma^2$ and $pF_{\alpha,p,\nu}\to\chi^2_{\alpha,p}$.) The results also apply to the fixed radius confidence sets of Section 2, for which it was proved that the coverage probability is bounded below by $1-\alpha$. For these confidence sets, it follows from Theorem 4.2 that the simultaneous intervals constructed will have confidence coefficient $1-\alpha$. Although these intervals will uniformly improve upon the Scheffé intervals in terms of coverage probability, they will, of course, have the same length.

Acknowledgement

This research was supported by National Science Foundation Grants No. DMS-8501973 and No. MCS80-03568 at Cornell University.

References

Berger, J. O. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* **8**:716-761.

Bock, M. E. (1982a). Minimax estimators that shift toward a hypershpere for mean vectors of spherically symmetric distributions. Technical Report #81-51, Department of Statistics, Purdue University.

Bock, M. E. (1982b). Employing vague inequality information in the estimation of normal mean vectors. (Estimators that shrink to closed convex polyhedra.) Technical Report #81-42, Department of Statistics, Purdue University.

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37**:1087-1136.

Casella, G. and Hwang, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**:688-698.

Casella, G. and Hwang, J. T. (1982). Decision-theoretic set estimation. Technical Report BU-773-M, Biometrics Unit, Cornell University.

Efron, B. and Morris, C. N. (1973). Stein's estimation rule and its competitors – and empirical Bayes approach. *J. Amer. Statist. Assoc.* **68**:117-130.

Faith, R. E. (1976). Minimax Bayes point and set estimators of a multivariate normal mean. Technical Report #66, Department of Statistics, University of Michigan.

Hwang, J. T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10**:868-881.

Hwang, J. T. and Casella, G. (1984). Improved set estimators for the mean of a multivariate normal distribution. *Statistics and Decisions*, Supplement Issue No. 1, 3-16.

Joshi, V. M. (1967). Inadmissibility of the usual confidence set for the mean of a multivariate normal population. *Ann. Math. Statist.* **38**:1868-1875.

Joshi, V. M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Ann. Math. Statist.* **40**:1042-1067.

Krishnaiah, P.R. (1965). Multivariate comparison tests in multi-response experiments, *Sankhyā, Series A*, 65-72.

Lindley, D. V. (1962). Discussion of Professor Stein's paper, "Confidence sets for the mean of a multivariate normal distribution." *J. Roy. Statist. Soc., Series B*, **24**:265-296.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78**:47-65.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley and Sons, Inc., New York.

TABLE 1

Values of a_0 for $q = 1$

p	α=.10	a=.05	p	a=.10	α=.05
4	.669	.633	15	9.888	9.675
5	1.517	1.458	16	10.731	10.510
6	2.392	2.281	17	11.574	11.345
7	3.211	3.083	18	12.418	12.182
8	4.036	3.893	19	13.263	13.020
9	4.865	4.710	20	14.109	13.859
10	5.697	5.531	21	14.955	14.699
11	6.532	6.355	22	15.802	15.539
12	7.369	7.182	23	16.650	16.381
13	8.207	8.011	24	17.498	17.223
14	9.047	8.842	25	18.346	18.065

TABLE 2

Coverage probabilities of the set C_{δ^L} with $a=a_0$, $\alpha=.10$

	p					
$\|\theta-\bar{\theta}\underset{\sim}{1}\|/\sigma$	4	8	12	16	20	24
0	.940	.991	.998	.999	.999	.999
1	.936	.990	.998	.999	.999	.999
2	.927	.985	.997	.999	.999	.999
3	.910	.976	.994	.998	.999	.999
4	.905	.956	.988	.997	.999	.999
6	.902	.930	.962	.983	.994	.998
8	.901	.918	.941	.962	.978	.989
10	.901	.912	.928	.946	.961	.974
20	.900	.903	.908	.914	.920	.927
50	.900	.901	.901	.902	.904	.905
100	.900	.900	.900	.901	.901	.901

TABLE 3

Coverage probabilities of the set C_{δ^L} with $a=p-3$, $\alpha=.10$

	p					
$\|\theta-\bar{\theta}\underset{\sim}{1}\|/\sigma$	4	8	12	16	20	24
0	.952	.995	.999	.999	.999	.999
1	.948	.994	.999	.999	.999	.999
2	.936	.990	.999	.999	.999	.999
3	.911	.982	.997	.998	.999	.999
4	.905	.958	.991	.998	.999	.999
6	.902	.931	.963	.985	.995	.998
8	.901	.919	.942	.964	.980	.990
10	.901	.912	.929	.947	.963	.978
20	.900	.903	.908	.914	.921	.928
50	.900	.901	.901	.902	.904	.905
100	.900	.900	.900	.901	.901	.901

TABLE 4

Coverage probabilities of the empirical Bayes confidence set $C^E_{X,\sigma}$, $\alpha=.10$.

	p					
$\lvert\theta-\bar{\theta}\underset{\sim}{1}\rvert/\sigma$	4	8	12	16	20	24
0	.937	.987	.997	.999	.999	.999
1	.934	.985	.996	.999	.999	.999
2	.920	.976	.993	.998	.999	.999
3	.892	.950	.984	.995	.998	.999
4	.891	.906	.941	.968	.991	.997
6	.896	.902	.921	.939	.952	.961
8	.898	.903	.916	.933	.950	.961
10	.898	.902	.912	.925	.938	.950
20	.899	.900	.903	.907	.912	.917
50	.899	.900	.900	.901	.902	.903
100	.900	.900	.900	.900	.900	.900

TABLE 5

Ratio of the radii of $C^E_{X,\sigma}$ to $C^0_{X,\sigma}$ for $\alpha=.10$

	p					
$\lVert X-\bar{x}\underset{\sim}{1}\rVert/\sigma$	4	8	12	16	20	24
0	.958	.883	.847	.823	.804	.789
1	.958	.883	.847	.823	.804	.789
2	.958	.883	.847	.823	.804	.789
3	.964	.883	.847	.823	.804	.789
4	.980	.907	.847	.823	.804	.789
6	.991	.964	.937	.906	.868	.818
8	.995	.980	.968	.955	.941	.926
10	.997	.988	.980	.973	.966	.959
20	.999	.997	.995	.994	.993	.991
50	.999	.999	.999	.999	.999	.999

TABLE 6

Coverage probabilities for the set $C^{E}_{X,s}$, $\alpha=.1$

	p=4				
	ν				
$\|\theta-\bar{\theta}\underset{\sim}{1}\|/\sigma$	2	5	10	20	30
0	.901	.909	.918	.926	.929
1	.901	.907	.916	.923	.926
2	.900	.903	.907	.912	.914
4	.900	.898	.897	.895	.894
6	.900	.899	.898	.897	.897
8	.900	.899	.899	.899	.899
10	.900	.899	.899	.899	.899
15	.900	.899	.900	.900	.900
20	.900	.899	.900	.900	.900

	p=10				
	ν				
$\|\theta-\bar{\theta}\underset{\sim}{1}\|/\sigma$	2	5	10	20	30
0	.903	.926	.953	.973	.981
1	.902	.924	.950	.971	.978
2	.902	.919	.942	.962	.971
4	.901	.909	.919	.927	.929
6	.900	.904	.908	.910	.910
8	.900	.902	.905	.907	.908
10	.900	.901	.903	.905	.905
15	.900	.901	.901	.902	.903
20	.900	.900	.901	.901	.902

TABLE 7

Ratio of the radii of $C^E_{X,s}$ to $C^0_{X,s}$, α=.1

p=4

$\|X-\bar{x}\underset{\sim}{1}\|/s$	ν = 2	5	10	20	30
0	.994	.980	.971	.965	.963
1	.994	.980	.971	.965	.963
2	.994	.980	.971	.965	.963
4	.994	.982	.981	.981	.980
6	.994	.992	.992	.991	.991
8	.996	.996	.995	.995	.995
10	.998	.997	.997	.997	.997
15	.999	.999	.999	.999	.999
20	.999	.999	.999	.999	.999

p=10

$\|X-\bar{x}\underset{\sim}{1}\|/s$	ν = 2	5	10	20	30
0	.983	.942	.913	.892	.884
1	.983	.942	.913	.892	.884
2	.983	.942	.913	.892	.884
4	.983	.942	.913	.892	.884
6	.983	.947	.946	.947	.948
8	.983	.971	.971	.972	.972
10	.984	.982	.982	.982	.983
15	.993	.992	.992	.992	.993
20	.996	.996	.996	.996	.996

TABLE 8

Coverage probabilities for the simultaneous intervals given in (4.13), $\alpha=.10$

	p=4				
	ν				
$[\Sigma n_i(\theta_i-\bar{\theta})^2/\sigma^2]^{\frac{1}{2}}$	2	5	10	20	30
0	.902	.912	.924	.933	.937
1	.902	.911	.923	.932	.936
2	.901	.906	.914	.921	.926
4	.899	.899	.898	.897	.896
6	.899	.899	.899	.899	.898
8	.900	.900	.899	.899	.899
10	.900	.900	.900	.900	.900
15	.900	.900	.900	.900	.900
20	.900	.900	.900	.900	.900

	p=10				
	ν				
$[\Sigma n_i(\theta_i-\bar{\theta})^2/\sigma^2]^{\frac{1}{2}}$	2	5	10	20	30
0	.904	.930	.959	.979	.986
1	.903	.928	.956	.977	.984
2	.902	.923	.948	.970	.978
4	.901	.911	.925	.935	.938
6	.901	.905	.911	.916	.907
8	.900	.903	.907	.910	.911
10	.900	.902	.904	.907	.908
15	.900	.901	.902	.903	.904
20	.900	.900	.901	.902	.902

TABLE 9

Ratio of the radii of the intervals in (4.13) to the corresponding Scheffé intervals

$[\Sigma n_i(\bar{X}-\bar{\bar{x}})^2/s^2]^{\frac{1}{2}}$	p=4, ν = 2	5	10	20	30
0	.992	.976	.966	.961	.958
1	.992	.976	.966	.961	.958
2	.992	.976	.966	.961	.958
4	.992	.984	.983	.983	.983
6	.994	.993	.993	.993	.993
8	.997	.996	.996	.996	.996
10	.998	.997	.997	.997	.997
15	.999	.999	.999	.999	.999
20	.999	.999	.999	.999	.999

$[\Sigma n_i(\bar{X}-\bar{\bar{x}})^2/s^2]^{\frac{1}{2}}$	p=10, ν = 2	5	10	20	30
0	.981	.937	.907	.886	.877
1	.981	.937	.907	.886	.877
2	.981	.937	.907	.886	.877
4	.981	.937	.907	.886	.877
6	.981	.949	.949	.951	.952
8	.981	.972	.973	.974	.975
10	.984	.982	.983	.984	.984
15	.993	.992	.992	.993	.993
20	.996	.996	.996	.996	.996